Modular Evolution of the Integrase Domain in the Ty3/Gypsy Class of LTR Retrotransposons

HARMIT S. MALIK AND THOMAS H. EICKBUSH*

Department of Biology, University of Rochester, Rochester, New York 14627-0211

A phylogenetic analysis of the Ty3/Gypsy group of retrotransposons identified a conserved domain (GPY/F) present in the integrases of several members of this group as well as of certain vertebrate retroviruses. The analysis suggested an evolutionary scheme for the acquisition and loss of the GPY/F domain as well as the acquisition of a chromodomain module in the integrase encoded by this group of elements that may direct targeting specificity in the host genome.

What directs the location of transposable elements' integration? A simple random insertion strategy is fraught with peril, as it may lead to both insertional mutations for the host and "dead" copies for the element. Natural selection will thus seek out members of a family of transposable elements that have hit upon strategies to maximize their own transmission frequencies (9, 19). One particular strategy may be for the transposable element to possess an insertion site preference in proximity to genes whose high levels of expression are guaranteed in the host genome but which do not disrupt expression of the genes themselves.

LTR (or long terminal repeat-bearing) retrotransposons can be divided into four major classes, the retroviruses, the Ty3/Gypsy group, the Ty1/Copia group, and the caulimoviruses. Each class contains a distinctive structural organization of open reading frames (ORFs) and enzymatic modules (29). Members of the Ty3/Gypsy group are of interest, because in sharp contrast to their neighbors, the retroviruses, they sometimes possess a remarkable site specificity of insertion. The *Saccharomyces cerevisiae* Ty3 element, for example, inserts within 5 bp of the start site of RNA polymerase III (Pol III)-transcribed genes (5). This turns out to be an excellent choice for insertion preference. Pol III-transcribed genes possess internal promoters; thus, the insertions are expected to have a minimal effect on gene expression levels (13). This remarkable preference is proposed to be brought about by the tethering of the Ty3 integration machinery to the Pol III transcription apparatus through protein-protein contacts with the transcription factors TFIII-B and -C (14). While many other members of the Ty3/Gypsy group have been sequenced, much less is known about their specificities (but see reference 7).

While the site specificity of Ty3 (and related retrotransposons) has attracted considerable biochemical and genetic effort, the evolutionary scheme by which this specificity was acquired or lost is still unknown (6). Previous attempts to determine the phylogeny of the Ty3/Gypsy group have been restricted by the lack of resolution afforded by the analyses of individual enzymatic domains (25, 28, 29). While increased resolution might be obtained by combining the information from all well-conserved domains, this could be done only if the elements have not swapped individual enzymatic domains. We investigated the possibility of domain swapping by comparing
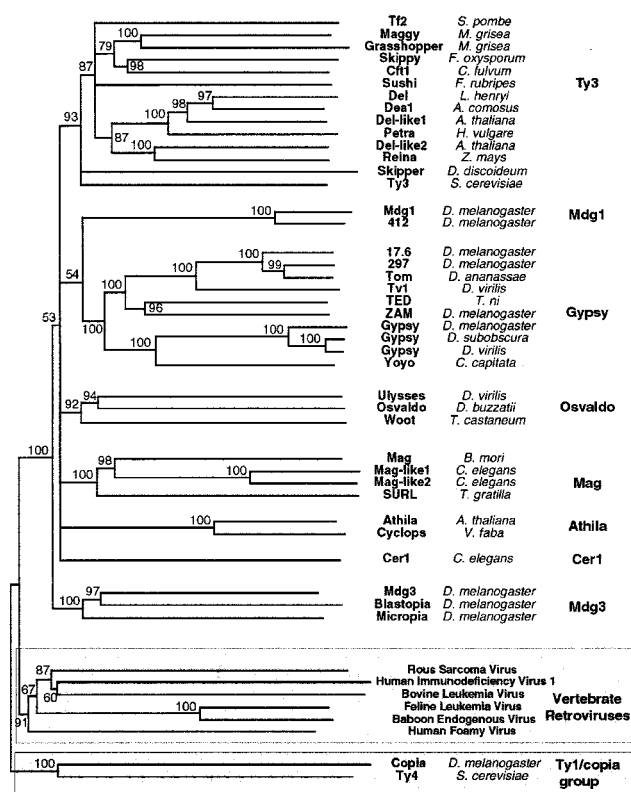

FIG. 1. Phylogenetic tree of the Ty3/Gypsy group and related classes of LTR retrotransposons. The tree is based on an alignment of the sum of the amino acids in the reverse transcriptase, RNase H, and integrase domains (approximately 700 amino acid positions) and is rooted based on the Ty1/Copia group (29). Selected vertebrate retroviruses representing the diversity of this group are included as members of a sister group to the Ty3/Gypsy group. Numbers adjoining the branches indicate bootstrap values (percentages from 1,000 trials) based on results of the neighbor-joining (22) method. The various retroelements and associated host genomes are indicated. Each major lineage of the tree has been named after one of its elements. *S. pombe, Schizosaccharomyces pombe; M. grisea, Magnoporthe grisea; F. oxysporum, Fusarium oxysporum; C. fulvum, Cladosporium fulvum; F. rubripes, Fugu rubripes; L. henryi, Lilium henryi; A. comosus, Ananus comosus; A. thaliana, Arabidopsis thaliana; H. vulgare, Hyphomicrobium vulgare; Z. mays, Zea mays; D. discoideum, Dictyostelium discoideum; D. ananassae, Drosophila ananassae; D. virilis, Drosophila virilis; T. ni, Trichoplusia ni; D. subobscura, Drosophila subobscura; C. capitata, Ceratitis capitata; T. castaneum, Tribolium castaneum; B. mori, Bombyx mori; C. elegans, Caenorhabditis elegans; T. gratilla, Tripneustes gratilla; V. faba, Vicia faba.*

* Corresponding author. Mailing address: Department of Biology, University of Rochester, Box 270211, Rochester, NY 14627-0211. Phone: (716) 275-7247. Fax: (716) 275-2070. E-mail: eick@uhura.cc.rochester.edu.

FIG. 2. Alignment of the GPY/F module. Coding regions from different elements starting from 40 residues downstream of the core integrase domain are aligned to highlight the conserved residues in a number of Ty3/Gypsy retrotransposons (upper group) and certain vertebrate retroviruses (lower group). The shading is to a 50% consensus (with MACBOXSHADE), with the black boxes indicating invariant amino acids and the gray boxes representing similar amino acids. The numbers in the alignment refer to the numbers of amino acids between the blocks of conserved residues. The thick horizontal line denotes the most conserved segment of this domain. Retrovirus abbreviations: Baboon ERV, baboon endogenous retrovirus; Feline ERV, feline endogenous regrovirus; Moloney MLV, Moloney murine leukemia virus; Gibbon LV, gibbon ape leukemia virus; Avian REV, avian reticuloendotheliosis virus; Walleye EHV, walleye epidermal hyperplasia virus; Snakehead RV, snakehead retrovirus.

the phylogenies of members of the Ty3/Gypsy group based on either the reverse transcriptase domain, the RNase H domain, or the core region of the integrase domain. Where sufficient resolution was afforded in the individual analyses (at least 50% bootstrap support in a neighbor-joining analysis [22]), no cases of disagreements among the three phylogenies were observed, suggesting the absence of any domain swapping (data not shown). We therefore combined all three data sets to obtain maximum resolution. Neighbor-joining analysis was carried out with PAUP* (version 4d64). This phylogeny is shown in Fig. 1 and has been rooted with the divergent Ty1/Copia retrotransposons as an outgroup.

Some surprising groupings that were not evident in previous analyses (25, 28, 29) emerged from this analysis of the Ty3/Gypsy group of elements. For example, the three lineages of Ty3/Gypsy retrotransposons that have been proposed to bear envelope domains, the well-characterized Gypsy group from *Drosophila melanogaster* (24), the plant errantiviruses Athila and Cyclops (28), and Osvaldo from *Drosophila buzzatii* (16), appear to be well separated. In addition, there is a clear grouping of Ty3 with Skipper, Tf2, and a number of additional plant and fungal retrotransposons.

The integrase domain of the Ty3/Gypsy group and retroviruses has been typically classified into three distinct subdomains (12). The N-terminal subdomain contains an HH-CC motif implicated in binding to LTR sequences but not in binding to target site DNA (12). The central core subdomain contains the catalytic D,D_{35}-E motif. Both these domains were used in the above-described phylogenetic analysis. The C-terminal subdomain of the integrase has been implicated in nonspecific binding to the target site in some cases (10); however, a recent study has also demonstrated the role of this subdomain in binding to the LTR of Ty3 elements (18). The C-terminal subdomain is the least well conserved of the three subdomains, showing great variation in sequence and length,

and is consequently not useful for a universal phylogenetic analysis.

We were intrigued by the possibility that this domain contains the factors necessary for the integration specificities of these elements. Thus, we turned our attention to the C-terminal subdomain. We were not surprised to find a high degree of conservation within this subdomain in the Ty3 lineage (Fig. 1), such subdomains being related by ancestry. More surprising was the finding that this subdomain is also well conserved among several additional members of the Ty3/Gypsy class of retrotransposons as well as certain vertebrate retroviruses (Fig. 2). The most conserved region of this module can be loosely identified as G-(D/E)-X_{10-20} (mostly hydrophobic residues)-K-L-X_2-(R/K)-(F/Y/W)-X-G-P-(F/Y)-X-(I/V), where the letters are in the single-letter amino acid codes and X refers to any amino acid. This module is hitherto referred to as GPY/F to highlight the best-conserved residues and for brevity. The GPY/F module is not universal in the Ty3/Gypsy group. Members of the Mag and Gypsy lineages and some in the Osvaldo lineage (Fig. 1) do not contain the GPY/F module. This situation is mirrored in the retroviral clade, where only a subset of retroviruses (indicated in Fig. 2) contains the GPY/F module. It is worth mentioning that all the retroviruses indicated in Fig. 2 that contain the GPY/F module group together phylogenetically (data not shown, but see reference 29). This differential retention of the GPY/F module is illustrated in the schematic shown in Fig. 3. It can be noted that two of the three clades that have lost the GPY/F module have acquired another ORF, which has been shown to be *env* (envelope)-like (16, 24).

It is striking that the integrase domain does not end with the GPY/F module in members of the Ty3 lineage (Fig. 3). In some of these elements, this portion of the ORF has been noted to have homology to a chromodomain (1, 15). The chromatin organization modifier domain (chromodomain) is a domain comprising about 50 amino acids that was originally iden-
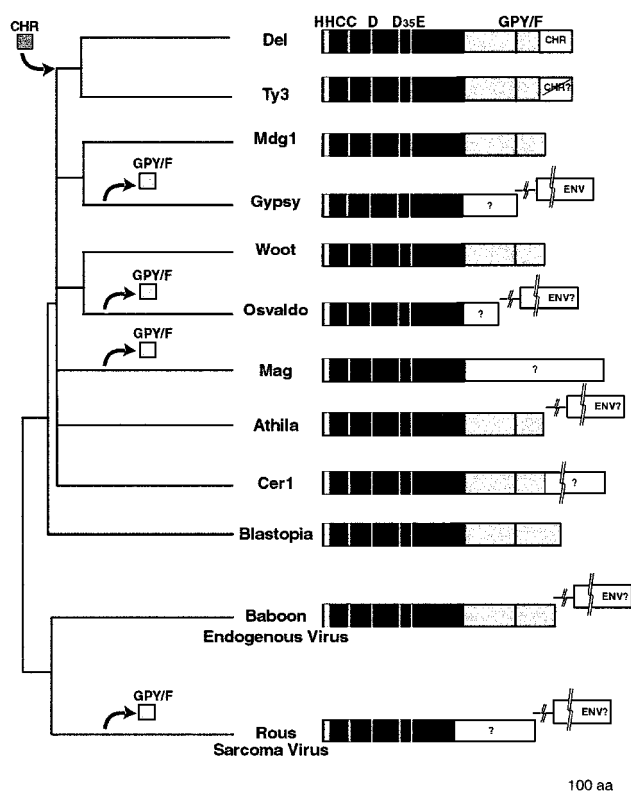
FIG. 3. Schematic evolution of the integrase domain in the Ty3/Gypsy group. The integrase domains of representatives from the Ty3/Gypsy groups identified in Fig. 1 are presented to the scale shown. The HH-CC and D,D$_{35}$-E motifs of the N-terminal and core integrase subdomains are highlighted (12). In addition, the GPY/F and chromodomain (CHR) modules of certain elements are illustrated. This hypothetical evolutionary scheme presents the three separate cases of loss of the GPY/F module as well as the monophyletic introduction of the chromodomain module in this group. Three apparently separate lineages of elements have also acquired a downstream *env* domain (16, 24, 28). Note that a similar loss of the GPY/F domain has occurred in vertebrate retroviruses. aa, amino acids.

tified as a protein sequence motif common to the *Drosophila* chromatin proteins Polycomb (Pc) and heterochromatin protein 1 (HP1) (20). Subsequently, these domains were identified in a variety of proteins that play a role in chromatin modification (1, 15), including factors that activate and repress transcription. Functional studies have shown that chromodomains are responsible for targeting the chromatin sites of action (17, 21). Recent nuclear magnetic resonance studies have also suggested that the chromodomain may mediate interactions between different proteins as a stand-alone protein module (3). We identified a chromodomain in all members of the Ty3 lineage (Fig. 1), except for the Ty3 element itself, using a combination of BLASTP and TBLASTN searches (2). In addition to the members of the Ty3 lineage represented in the phylogenetic diagram, we could identify chromodomains in four additional members that could not be included in Fig. 1 as their sequences were incomplete. An alignment of these chromodomain motifs with CLUSTAL W (27) is presented in Fig. 4. Much like other authentic-chromodomain-containing proteins, some retrotransposons match the universal chromodomain consensus better than others. However, all members of this lineage, except the Ty3 element itself, bear a large proportion of the critical (conserved) amino acid residues. When present in the Ty3 lineage, chromodomains are found at the

C-terminal end of the integrase domain, never as a stand-alone ORF, as previously noted (15). It is interesting that while Ty3 does not bear a chromodomain, it does bear a module of approximately the same size. This suggests a model for Ty3 in which an ancestral chromodomain module may have been replaced or become specialized. This specialization in Ty3 has been proposed to reflect the selection imposed by the haploid-diploid life cycle of *S. cerevisiae*, where any gene disruption is expected to have the most severe consequence (4, 13).

We can thus propose a scheme for the segmental evolution of the integrase domain in the Ty3/Gypsy group, summarized in Fig. 3. The common ancestor of the Ty3/Gypsy group contained only a GPY/F domain. This was supplemented with a chromodomain (or analog) in the Ty3 group, by an envelope domain in the plant errantiviruses, and by a still unknown domain in Cer1. This domain has been eliminated (replaced) on potentially three independent occasions with the evolution of the Gypsy, Mag, and Osvaldo lineages. Considerable biochemical attention has been paid to the integration specificities in the Ty3/Gypsy group (7, 14, 26). In a recent study, a chimeric Ty3/Moloney murine leukemia virus integrase was created, with the C-terminal end of the virus being replaced with that of Ty3 (8). While this chimeric integrase was functional, it lacked the specificity of Ty3 for tRNA genes in human cell lines. Unfortunately, a caveat in interpreting this result is that if Ty3 has indeed specialized its interaction module, this specificity may not be observed outside the yeast genome. We are confident that future efforts based on our analysis of these domains will help streamline efforts to elucidate the evolution of different integration strategies in the Ty3/Gypsy group and perhaps the vertebrate retroviruses as well. The presence of chromodomains in other members of the Ty3 lineage may implicate this module in directing targeting to sites in the host genome via protein-protein interactions (with authentic-chromodomain-containing proteins serving as an analogy). However, this specificity may be loosely defined relative to that of the Ty3 element itself. Thus, it is likely that other members of the Ty3 lineage are specific for locations in or about expressed genes, for example, and only examination of numerous copies will elucidate the region-specific targeting that is brought about by the chromodomain. It is possible that the newly defined GPY/F domain may also be involved in some (as yet unknown) specificity that allows directed integration into safe havens and, thus, long-term maintenance of the elements in the host genome. The presence of *env* domains in retroviruses of both vertebrates and insects may obviate the need for such a domain; although some strong site preference has been found in members of the Gypsy lineage (see reference 7 for instance), the biochemical basis for which remains unknown.

Acquisition of target specificity based on a chromodomain is a novel means by which a transposable element can direct integration efforts in a host genome by relying on protein-protein interactions, making it independent of the actual DNA sequence that is the target. The alternative would be to encode a site-specific endonuclease, as is the case with group II introns (23) and some non-LTR retrotransposable elements (30). It is perhaps not surprising that this domain inclusion usually occurs in the Ty3/Gypsy group at the end of the integrase ORF, where it probably has a minimal impact on the core set of enzymatic activities of the element. Similar strategies exist not only in vertebrate retroviruses but also in the Ty1/Copia group of retrotransposons. In the latter group, targeting of Ty5 to silenced regions of the *S. cerevisiae* chromosome was shown to be abolished by a single amino acid change in a domain downstream of the core integrase domain (11).

```
D. melanogaster  Polycomb     Y A A E K I I Q K R - V K K G V V E - - - - Y R V K W K G W N Q Y - - - R N T W E P E V N I
D. melanogaster  Su(Var) 3-9  Y V V E R I E C V E M E D Q Y Q P V - - - - F F V K W L G Y H D S - - - E N T W E S L A N V
H. sapiens       HP1          Y V V E K V L R R R - V V K G Q V E - - - - Y L L K W K G F S E E - - - H N T W E P E K N L
M. musculus      MOD3         F A A E C I L S K R - L R K G K L E - - - - Y L V K W R G W N S S K - - - H N S W E P E E N I
S. pombe         SWI6         Y V V E K V L K H R M A R K G G G Y E - - - Y L L K W E G Y D D P S - - D N T W S S E A D C


C. fulvum        Cft1         F E V E K I L D K - - - - K G Q R - - - - - Y L V K W K G Y D E S - - - E N T W E P R I N L
F. oxysporum     Skippy       Y E A E A I R D T R - K I N G Q R E - - - - Y L I K W K N Y P E N - - - E N T W E P P K H L
M. grisea        Maggy        Y E V E E I L D S F W E T R G R G G R R L K Y I V R W A G Y S E P - - - - - T T E P A D Y L
M. grisea        Grh          W A V E A I L A A K - N R R G R G G G R Q - V L V K W Q G Y D N P - - - - - T W E P L E L M
L. henryi        Del          E K P V Q V L A S E - S K V L R N K I I L M V K V L W Q H H S E E - - - E A T W E L E A D M
A. thaliana      Del-like 1   A R P V R V L E R R - I K E L R K K I P L I K V L W D C D G V - - - E E T W E S E T R M
A. thaliana      Del-like 2   K A P E T I L E R K - M V N R Q G R A A T K V L V K W V N E P V E - - - E A T W E F L F D L
A. thaliana      Del-like 3   A W P V R I M D Q M - K K G T R G K S M D L L K I L W N C G G R E - - - E Y T W E T E T K M
H. vulgare       Petra        E Y L V C I L D Q A - E R T N R R R N L K F L K V Q W S N H S E K - - - E A T W E R E D R L
Z. mays          Reina        K I P T E V L E S R - L L R K G N K V I P Q L L I R W S N W P A S - - - L S T W E D E H A I
Z. mays          Rle          Q V P E T I L A T R - L R R A G K V V K Q L L I K W S S W H S S - - - L A T W E D E A D I
A. capillus      ARET-1       I L V E Q V L A H K - D R K V I G Q V A R R Y L V K F K N Y P P M - - - D A K W M E E H E V
N. alata         Tna1-2       R R F L A I I D M Q - V R K L R N K E I A S V K V L W R N Q Q V E - - - E A T W E E E D E M
S. pombe         Tf2          S D I G T I L H I L - E H K N R E Q - V L Y L N V K Y I S N L N P S T I M S G W T T L A T A
F. rubripes      Sushi        W S V N K L L A R R - - R R G R G F Q - - - Y L V D W V G Y G P E - - - D Q S W V P D S Y L
D. discoideum    Skipper      Y E V E S I L N H D K V K K M - - - - - - - Y L V K F K G Y P - - - - - E P E W I R E V D T


S. cerevisiae    Ty3          I G I D T T H K T Y - L C H M Q - - - - D V D P T L S V E Y S E A - - - E F C Q I P E R T R
```

FIG. 4. Alignment of putative chromodomains from various retrotransposons. The first block includes representatives of authentic chromodomains from a variety of chromatin-modifying proteins: *D. melanogaster* Polycomb and Su(Var) 3-9, *Homo sapiens* HP1, *Mus musculus* MOD3, and *Schizosaccharomyces pombe* SWI6. EMBL accession numbers for the elements in Fig. 4 which had to be conceptually translated from their DNA sequences (and are not available in protein databases) are as follows: M77661 for *grasshopper* or *grh*, AC002534, AB007727, and Z97342 for *Del-like 1 to 3* (also called Tma), Y14573 for *Petra*, U69258 for *Reina*, AF057037 for *Reina-like element* (Rle), AB003364 for ARET-1, and Z35426 for Tna1-2. EMBL accession numbers for the elements in Fig. 4 which had to be conceptually translated from their DNA sequences (and are not available in protein databases) are as follows: M77661 for *grasshopper* or *grh*, AC002534, AB007727, and Z97342 for *Del-like 1 to 3* (also called Tma), Y14573 for *Petra*, U69258 for *Reina*, AF057037 for *Reina-like element* (Rle), AB003364 for ARET-1, and Z35426 for Tna1-2. Del-like 1 to 3, ARET-1, and Rle are not represented in Fig. 1. The sequence of the Dea1 element (Fig. 1) is not included, as the sequence of the entire chromodomain module is not available. The shading scheme is similar to that presented in Fig. 2. *A. capillus, Adiantum capillus*; *N. alata, Nicotiana alata*. (All other genera are as defined in the legend to Fig. 1.)

Nucleotide sequence accession numbers. The alignments used in Fig. 1 have been deposited in the EMBL online database (9a) under accession no. DS36733 (reverse transcriptase), DS36732 (RNase H), and DS36734 (integrase).

REFERENCES

1. **Aasland, R., and A. F. Stewart.** 1995. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. Nucleic Acids Res. **23:**3168–3173.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Ball, L. J., N. V. Murzina, R. W. Broadhurst, A. R. Raine, F. J. Stott, A. G. Murzin, P. B. Singh, P. J. Domaille, and E. D. Laue.** 1997. Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. EMBO J. **16:**2473–2481.
4. **Bushman, F.** 1995. Targeting retroviral integration. Science **267:**1443–1444.
5. **Chalker, D. L., and S. B. Sandmeyer.** 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. **6:**117–128.
6. **Craigie, R.** 1992. Hotspots and warm spots: integration specificity of retroelements. Trends Genet. **8:**187–190.
7. **Dej, K. J., T. Gerasimova, V. G. Corces, and J. D. Boeke.** 1998. A hotspot for the Drosophila Gypsy retroelement in the ovo locus. Nucleic Acids Res. **26:**4019–4025.
8. **Dildine, S. L., J. Respess, D. Jolly, and S. B. Sandmeyer.** 1998. A chimeric Ty3/Moloney murine leukemia virus integrase protein is active in vivo. J. Virol. **72:**4297–4307.
9. **Doolittle, W. F., and C. Sapienza.** 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature **284:**601–603.
9a. **EMBL Database Website.** Modified 16 March 1999. [Online.] http://www.e-bi.ac.uk. [30 March 1999, last date accessed.]
10. **Engelman, A., A. B. Hickman, and R. Craigie.** 1994. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J. Virol. **68:**5911–5917.
11. **Gai, X., and D. F. Voytas.** 1998. A single amino acid change in the yeast Ty5 abolishes targeting to silent chromatin. Mol. Cell **1:**1051–1055.
12. **Khan, E., J. P. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. **19:**851–860.
13. **Kinsey, P. T., and S. B. Sandmeyer.** 1995. Ty3 transposes in mating populations of yeast: a novel transposition assay for Ty3. Genetics **139:**81–94.
14. **Kirchner, J., C. M. Connolly, and S. B. Sandmeyer.** 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science **267:**1488–1491.
15. **Koonin, E. V., S. Zhou, and J. C. Lucchesi.** 1995. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. Nucleic Acids Res. **23:**4229–4233.
16. **Labrador, M., and A. Fontdevila.** 1994. High transposition rates of Osvaldo, a new *Drosophila buzzatii* retrotransposon. Mol. Gen. Genet. **245:**661–674.
17. **Messmer, S., A. Franke, and R. Paro.** 1992. Analysis of the functional role of the Polycomb chromo domain in *Drosophila melanogaster*. Genes Dev. **6:**1241–1254.
18. **Nymark-McMahon, M. H., and S. B. Sandmeyer.** 1999. Mutations in nonconserved domains of Ty3 integrase affect multiple stages of the Ty3 life cycle. J. Virol. **73:**453–465.
19. **Orgel, L. E., and F. H. Crick.** 1980. Selfish DNA: the ultimate parasite. Nature **284:**604–607.
20. **Paro, R., and D. S. Hogness.** 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc. Natl. Acad. Sci. USA **88:**263–267.
21. **Platero, J. S., T. Hartnett, and J. C. Eissenberg.** 1995. Functional analysis of the chromo domain of HP1. EMBO J. **14:**3977–3986.
22. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
23. **Shub, D. A., H. Goodrich-Blair, and S. R. Eddy.** 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. Trends Biochem. Sci. **19:**402–404.
24. **Song, S. U., T. Gerasimova, M. Kurkulos, J. D. Boeke, and V. G. Corces.** 1994. An env-like protein encoded by a Drosophila retroelement: evidence that Gypsy is an infectious retrovirus. Genes Dev. **8:**2046–2057.
25. **Springer, M. S., and R. J. Britten.** 1993. Phylogenetic relationships of reverse transcriptase and RNase H sequences and aspects of genome structure in the Gypsy group of retrotransposons. Mol. Biol. Evol. **10:**1370–1379.
26. **Steele, S. J. S., and H. L. Levin.** 1998. A map of interactions between the proteins of a retrotransposon. J. Virol. **72:**9318–9322.

27. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.

28. **Wright, D. A., and D. F. Voytas.** 1998. Potential retroviruses in plants. Tat1 is related to a group of *Arabidopsis thaliana* Ty3/Gypsy retrotransposons that encode envelope-like proteins. Genetics **149:**703–715.

29. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.

30. **Yang, J., H. S. Malik, and T. H. Eickbush.** The endonucleases encoded by site-specific non-LTR retrotransposons are related to restriction enzymes. Submitted for publication.